8.3 Describing and Analyzing Data

Goals for the Day

1. Measures of Center
2. Measures of Spread
3. Using your Calculator
4. Empirical Rule

1

Measures of Center

Mean (Average Value)

- Simple, arithmetic average of the data.
  - Sum all numbers and divide by the sample size (n).
- Same calculation for the population and sample mean (just different notation).
  - Sample mean = $\bar{x}$ (pronounced "x-bar")
  - Population mean = μ (Greek letter mu)
- Mean is NOT a resistant measure.
  - This means it is heavily affected by outliers.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Example 1 – Mean

Data: 1, 5, 2, 9, 3

$$\bar{x} = \frac{1 + 5 + 2 + 9 + 3}{5} = 4$$

Now change 3 to 30

New $\bar{x} = 9.4$

Median (Middle Value)

- The <u>middle</u> value in an ordered list.
- Median IS a resistant measure.
  - NOT affected by outliers.

Example 2 – Median

<u>Case 1 – Odd *n*</u>

7 Obs: 10, 5, 6, 1, 3, 9, 8

Sorted: $\cancel{1}, \cancel{3}, \cancel{5}, 6, \cancel{8}, \cancel{9}, \cancel{10}$

$$Med = 6$$

<u>Case 2 – Even *n*</u>

8 Obs: 10, 5, 6, 1, 3, 9, 8, 3

Sorted: $\cancel{1}, \cancel{3}, \cancel{3}, 5, 6, \cancel{8}, \cancel{9}, \cancel{10}$

$$Med = \frac{5+6}{2} = 5.5$$

Mode (Most Common Value)

- The most frequently occurring value(s).
  - Unimodal data has one mode.
  - Bimodal data has 2 modes.
  - Multimodal data has more than 2 modes.
  - Can be no modes (every value is distinct).
- This is the only measure of center that can be used with categorical data.
  - Ex) Most common favorite color (can't average this)

Example 2 – Median

Case 2

8 Obs: 10, 5, 6, 1, 3, 9, 8, 3

Mode = 3 (twice)

2

Measures of Spread

Range

- Range = Max - Min
- Gives idea of the entire ”range” of values, how much distance do they span in total.
- Ex) Case 2: Range = 10 – 1 = 9

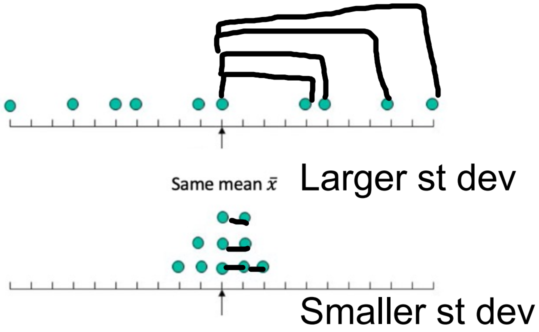
Example 2 – Median

<u>Case 2</u>

8 Obs: 10, 5, 6, 1, 3, 9, 8, 3

Standard Deviation

- Complex formula that measures the average distance that each data point is from the mean.



Sample Standard Deviation = S = $\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

Population Standard Deviation = $\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$

(Greek letter sigma)

3

Using your Calculator

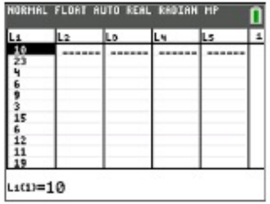
Using Your Calculator

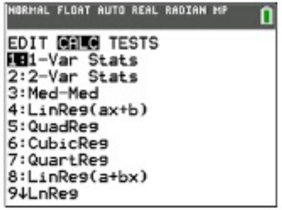
Steps for the TI-83/84	Steps for the TI-30XS MultiView	Steps for the TI-30 XIIS
1. Enter data: STAT → Edit → Enter data in L_1 *(Demo dataset: 10, 23, 4, 6, 9, 3, 15, 6)* 2. Calculate: STAT → CALC → 1-Var Stats a) List is L_1. b) Leave FreqList blank. c) Calculate!	1. Data → Enter data in L_1 2. 2^{nd} → stat → 1-Var Stats a) DATA: L1 b) FRQ: ONE c) CALC	1. 2^{nd} → STAT → 1-VAR (Enter) 2. DATA X_1 = # (scroll down) FRQ = 1 (for ALL Xs, scroll down) X_2 = # (scroll down) … 3. STATVAR (scroll across) 4. To exit this menu: 2^{nd} → EXIT STAT → Y

Inputs

Data here:

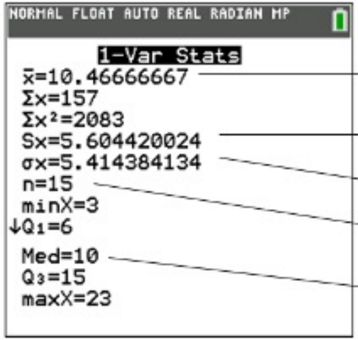
10, 23, 4, 6, 9,
3, 15, 6, 12, 11,
19, 10, 6, 8, 15





Results

Does not give median



Mean
Sample st dev
Population st dev
Sample size
Median

Using Your Calculator Example

Example 3: Find the mean, median, mode and sample standard deviation of the following dataset.

- Data (7 obs): 35, 70, 31, 37, 65, 38, 38

<u>Results</u>

Mean $\bar{x} = 44.86$

Sample st dev: $S_x = 15.72$

Pop st dev $\sigma_x = 14.55$

Med = 38

Other Considerations

3

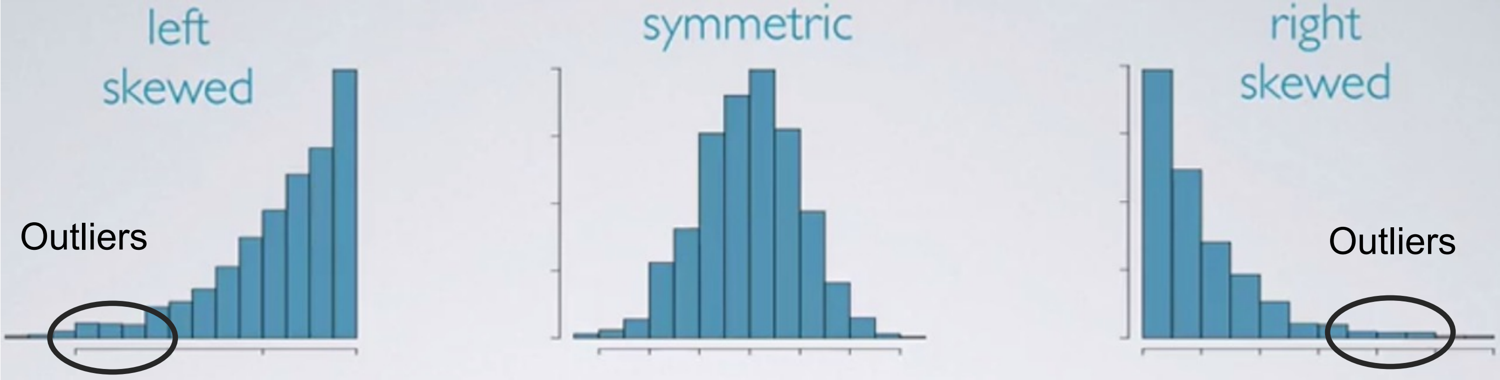
- Outliers
  - Data values that are extreme when compared to the rest of the data.
  - Can significantly impact measures of center and spread.

Example:
- Data (7 obs): 31, 35, 37, 38, 38, 65, 70

Outliers

Other Considerations

Types of Distributions



Best measure of center: Median | Mean | Median

4

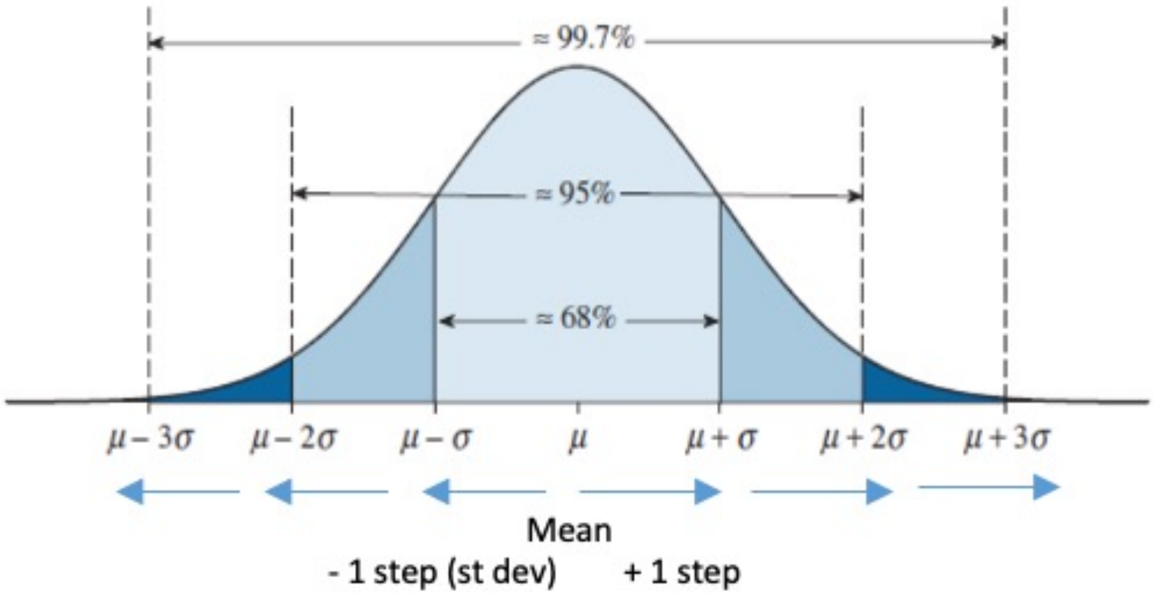
Empirical Rule

Empirical Rule (68 – 95 – 99.7 Rule)

- When data is approximately bell shaped, the standard deviation allows us to make fairly accurate approximations about the locations of our data values.

  - 68% of the data lies within 1 standard deviation of the mean.
  - 95% of the data lies within 2 standard deviations of the mean.
  - 99.7% of the data lies within 3 standard deviations of the mean.

Empirical Rule (68 – 95 – 99.7 Rule)



- We can use these breakdowns to find probabilities within certain intervals.

Empirical Rule Examples

Example 5: Suppose that IQ scores have a bell-shaped distribution with a mean of 105 and a standard deviation of 15. Using the empirical rule answer the following questions:

a) What percentage of IQ scores are greater than 75?

97.5%

Step 1
Draw and label curve

Step 2
Shade area of interest

b) b) Between which two values do the middle 68% of IQ scores fall between?

(90, 120)